

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 13 日現在

機関番号：14701

研究種目：挑戦的萌芽研究

研究期間：2012～2013

課題番号：24650085

研究課題名(和文)機能性音声のデザイン方法論の研究

研究課題名(英文) Investigations on design methodology of functional voices

研究代表者

河原 英紀 (KAWAHARA, Hideki)

和歌山大学・システム工学部・教授

研究者番号：40294300

交付決定額(研究期間全体)：(直接経費) 3,000,000円、(間接経費) 900,000円

研究成果の概要(和文)：理解度の向上、好感度の向上、警戒心の喚起、安心感の醸成等、言語以外の音声そのものが持つ属性をデザインするための方法論を構築することを目的として研究を進めた。そのために、音声研究のデファクトスタンダードとなっている代表者の発明になる音声分析変換合成システムSTRAIGHTを拡張し、有用なツール群を生み出した。そのためにSTRAIGHTの分析に用いられる基本周波数抽出法を高速・高時間分解能・高精度なものに拡張した。これらの過程で、STRAIGHTに基づくモーフィングが、任意の数の音声の様々な物理属性を個別にかつ時間毎に自由に操作できるものとして本質的に拡張され、本格的探索的研究の可能性が生まれた。

研究成果の概要(英文)：Aiming at establishing a design methodology of voice attributes to facilitate understanding, affection, alert and safety feeling by non-linguistic voice aspects, this research project was conducted. It was based on a de facto standard speech research framework STRAIGHT, developed by the prime investigator and yielded several useful research tools. A highly accurate and computationally efficient F0 extractor with higher temporal resolution than conventional methods is an example of such extension and development. In this course of investigations, a revolutionary new morphing framework was invented. The new method is capable of morphing arbitrary many voice examples in one shot procedure with individual control of each physical speech parameter and enabled time dependent weight setting. This new morphing method and other research products of the project provides basis of the substantially new research project.

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：音声合成 音声知覚 音声コミュニケーション 音声インタフェース 非言語

1. 研究開始当初の背景

(1)本課題を企画した2011年の中盤において、音声認識、音声合成技術は、計算機能力の急速な向上と大量の音声および言語データの蓄積により、日常生活に入ることが可能なレベルに達し、本格的な普及の時代に入りつつあった。しかし、それらは文字として表すことのできる一部の言語情報に限られており、人間の音声の有する非言語情報およびパラ言語情報の利用は、研究レベルにおいても大きく立ち後れている状況であった。

(2)研究代表者が発明した高品質な音声分析・変換・合成を可能とする STRAIGHT は、音声研究のデファクトスタンダードとして、広範な分野で用いられるようになっていた。特に、この方法が可能にした音声モーフィングは、従来では困難であった人間の音声の有する非言語情報およびパラ言語情報の強力な研究手段となる見通しが得られるようになりつつあった。

2. 研究の目的

自然な文字である手書き文字よりも、適切にデザインされた活字は遥かに読み易い。音声言語も、適切にデザインすることにより、自然な肉声そのままのものよりも、明瞭に聴き取ることができ、内容を良く理解し記憶でき、より深い感動を得、心の平安を得、警戒心を喚起できるなど、用途に応じたものとしてすることができる可能性がある。本研究課題は、最新の音声変換技術、合成技術を利用することにより、そのような用途に応じた機能を有する音声をデザインするための方法論を明らかにすることを目的とする。

3. 研究の方法

(1)研究の基盤としては、申請者が開発しており、現在、音声知覚および音声合成研究のデファクトスタンダードとなっている STRAIGHT を必要に応じて拡張して用いる。

(2)卓越した音声を収集したデータベース等を用い、強い印象を与える力を有する音声素材を選択する。

(3)STRAIGHT を拡張した高度な分析手法を用いて、選択した音声素材を分析し、強い印象の基盤となっている音声の知覚特性ならびに物理特性を明らかにする。

(4)STRAIGHT およびその発展形である TANDEM-STRAIGHT に基づく音声モーフィングを高度な分析方法に基づいて拡張するとともに、分析事例を蓄積し、機能的音声デザインの方法論を探索する。

4. 研究成果

(1)初年度は、研究環境としてのツール群の整備と手法の構築、対話的操作環境の構築を開始した。この中で、まず大量の音声素材を分析して用いる音声素材を選択する過程において、必要となる高精度かつ高速の基本周波数分析法を発明し、さらに、翌年度において原理を高度化するとともに、統計的手法に基づく初期推定値の頑健化を行うことにより、基本周波数の高精度かつ高時間分解能の

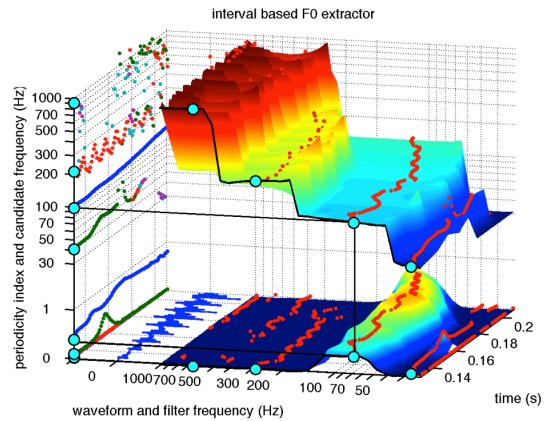


図 1 高速高精度の基本周波数抽出

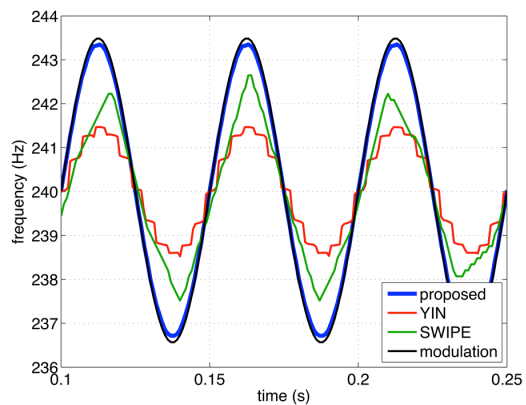


図 2 基本周波数抽出の比較

計測に用いることのできる方法が生み出されるに至った。

図 1 に、この方法における初期推定の概要を示す。下の曲面に示される波形の対称性に基づく指標のピークに基づいて、上の曲面から基本周波数の候補が選択されている。図 2 に、この方法と従来の方法の比較例を示す。黒線で示されている基本周波数の真値に、提案した方法が正確に追従していることが分る。

(2)中核的研究手段である STRAIGHT の処理内容を見直し、合成部分の実装を時変フィルタとして明確化し数理的基盤を整備した。この改造に際し、合成部分を汎用性の高い関数ポインタを用いた構造とした。その結果、合成の音源として、従来のパルス列だけではなく整数比関係にある正弦波の集合や、さらには非整数比の正弦波の集合を用意に用いることを可能にした。この STRAIGHT の再構成は、音声知覚研究のためのツールとしての有用性を向上させただけでなく、後述のモーフィングの本質的な拡張の基盤となった。

(3)合成部分の時変フィルタとしての再構築を応用し、音声の言語情報と任意の非音声信号をクロス合成する機構を開発した。本課題の成果の応用例である。

(4)研究用ツールの整備の一環として、聴覚心理実験用の新しい対話的インタフェースを有するソフトウェアを開発した。これは、これまでの統計的方法を見直し良く包含することができる一般化線形モデルに適した

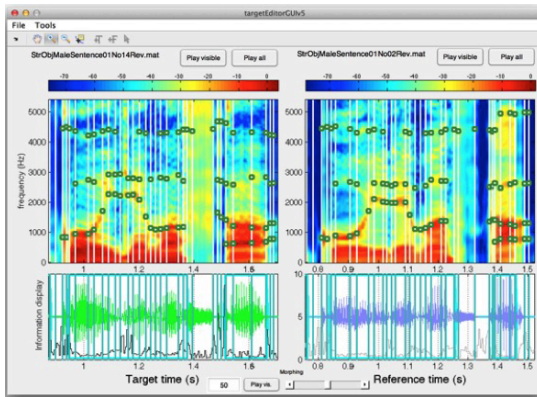


図 3 新しいモーフィング GUI

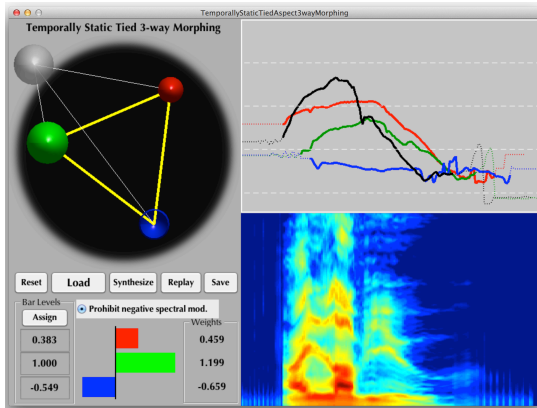


図 4 3 事例モーフィングの例

実験用ユーザインタフェースを組込んだものであり、聴覚心理実験の遂行と結果の分析を効率よく進めることを可能とした。

(5) STRAIGHT に基づく処理の高品質化の障害となっていた、非周期性指標の合理的分析を可能にする、新しい群遅延の安定化手法を明らかにした。本課題の中では、本格的な応用に至っていないが、STRAIGHT を本質的に再構築する可能性が示されたことは、大きな成果である。

(6) 本課題の二年目の 4 月から 5 月にかけて、②の論文をまとめるために、これまで研究協力を進めていた Schweinberger 教授（独、イエナ大学）の招きで研究室に短期滞在した際に、STRAIGHT に基づくモーフィングを本質的に拡張するアイデアを得た。これは、従来は 2 つの事例間においてのみ定式化されていたモーフィングを、任意の事例間に拡張するものであり、併せて、個別の属性を時系列の重みにより独立に制御することを、外挿の場合にも破綻しない形で可能にするものである。

これは当初想定していなかった新たな枠組みでの本格的な探索的研究を可能にするものである。この可能性を追求するために、新規にユーザインタフェースを再設計し、具体的な GUI として実装することまで進めることができた。

図 3 にモーフィングのために必要な事例間の対応情報を設定するための GUI を示し、図 4

に、3 事例間のモーフィングの応用例の操作 GUI を示す。

(7) 基本周波数抽出法の高度化による実現された高い時間分解能と、時変フィルタを組み合わせることにより、強い印象を与える歌唱技法の印象の操作を実現した。

5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 2 件）

① Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G. and Zäske, R., Speaker perception, WIREs Cognitive Science, 5(1), pp.15-25, 2014. 査読有り DOI:10.1002/wcs.1261

② Taiki Nishi, Ryuichi Nisimura, Toshio Irino and Hideki Kawahara, Controlling linguistic information and filtered sound identity for a new cross-synthesis vocoder, Acoustical Science and Technology, 34(4), pp.287-288, 2013. 査読有り DOI:10.1250/ast.34.287

〔学会発表〕（計 14 件）

① 坂口 諒, 小林 真優子, 入野 俊夫, 西村 竜一, 河原 英紀, 日本音響学会春季研究発表会, 1-R5-22, 東京, 2014. 3. 11.

② 吉元 照貴, 西村 竜一, 入野 俊夫, 河原 英紀, 外挿が可能な時変多属性任意事例数モーフィングを用いた文章音声好感度の改善について, 日本音響学会春季研究発表会, 1-R5-22, 東京, 2014. 3. 10.

③ 吉元 照貴, 西村 竜一, 入野 俊夫, 河原 英紀, 文章音声の好感度評価と属性別モーフィングを用いた要因の検討について, 電子情報通信学会音声研究会, 名古屋, 2014. 1. 23.

④ M. Sakaguchi, M. Kobayashi, R. Nisimura, T. Irino, H. Kawahara, Spectrally estimated vocal tract lengths of singing voices and their contributing factors, MAVEBA 2013, Firenze Italy, 2013. 12. 17. 査読有り

⑤ H. Kawahara, M. Morise, K. Sakakibara, Temporally fine F0 extractor applied for frequency modulation power spectral analysis of singing voices, MAVEBA 2013, Firenze Italy, 2013. 12. 17. 査読有り

⑥ 吉元 照貴, 西村 竜一, 入野 俊夫, 河原 英紀, 基本周波数操作による音声の好感度改善に関連する物理的特徴の検討, 日本音響学会秋季研究発表会, 1-P-11c, 豊橋, 2013. 9. 25.

⑦ Yuri Nishigaki, Ken-Ichi Sakakibara, Masanori Morise, Ryuichi Nisimura, Toshio Irino and Hideki Kawahara, Controlling “shout” expression in a Japanese POP singing performance: analysis and suppression study, Interspeech2013, Lyon France, 2013. 8. 28. 査読有り

⑧ Hideki Kawahara, Masanori Morise and Ken-Ichi Sakakibara, Interference-free observation of temporal and spectral features in "shout" singing voices and their perceptual roles, SMAC-SMC 2013, Stockholm Sweden, 2013. 8.1. 査読有り

⑨ Taiki Nishi, Ryuichi Nisimura, Toshio Irino, Hideki Kawahara, Modulation transfer function design for a flexible cross synthesis VOCODER based on F0 adaptive spectral envelope recovery, APSIPA ASC 2012, Hollywood, CA, USA, 2012.12.6. 査読有り

⑩ Hideki Kawahara, Masanori Morise, Ryuichi Nisimura and Toshio Irino, An interference-free representation of group delay for periodic signals, APSIPA ASC 2012, Hollywood, CA, USA, 2012.12.6. 査読有り

⑪ Hideki Kawahara, Masanori Morise, Ryuichi Nisimura, Toshio Irino, Deviation measure of waveform symmetry and its application to high-speed and temporally-fine F0 extraction for vocal sound texture manipulation, Interspeech2012, Portland, OR, USA, 2012.9.10, 査読有り

⑫ Josh H. McDermott, Daniel P. W. Ellis and Hideki Kawahara, Inharmonic Speech: A Tool for the Study of Speech Perception and Separation, SAPA-Scale Conference 2012, Portland, OR, USA, 2012.9.9, 査読有り

⑬ 河原 英紀, 森勢 将雅, 西村 竜一, 入野 俊夫, 音源およびスペクトル包絡の時間的微小構造の加工と歌唱音声の印象への影響について, 情報処理学会第95回音楽情報科学研究会, 東京, 2012.6.2

⑭ 西 大輝, 西村 竜一, 入野 俊夫, 河原 英紀, 楽器音や動物の鳴声の音色と音声の言語情報を保持したクロス合成 VOCODER, 情報処理学会第95回音楽情報科学研究会, 東京, 2012.6.2

[その他]

ホームページ等

6. 研究組織

(1) 研究代表者

河原 英紀 (KAWAHARA, Hideki)

和歌山大学・システム工学部・教授

研究者番号：40294300

(2) 研究分担者

森勢 将雅 (MORISE, Masanobu)

山梨大学・大学院医学工学総合研究部・助教

研究者番号：60510013